

W. Jassem, G. Demenko

ZALEŻNOŚĆ PRZEBIEGU PARAMETRU F_0
OD DŁUGOŚCI FRAZY
I DZWIĘCZNOŚCI SEGMENTALNEJ

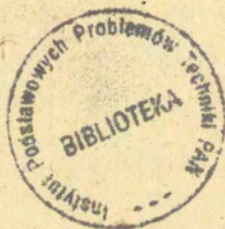
29/1989

P. 269



WARSZAWA 1989

Praca wpłynęła do Redakcji dnia 30 listopada 1988 r.



56741



Na prawach rękopisu

Instytut Podstawowych Problemów Techniki PAN

Nakład 140 egz. Ark.wyd. 2,4 Ark.druk. 3

Oddano do drukarni w kwietniu 1989 r.

Nr zamówienia 258/89

Warszawska Drukarnia Naukowa, Warszawa,
ul.Śniadeckich 8

WIKTOR JASSEM
GRAŻYNA DEMENKO

Zakład Fonetyki Akustycznej
Pracownia Analizy Mowy
IPPT PAN

Zależność przebiegu parametru F_0 od długości frazy i
dźwięczności segmentalnej

Streszczenie

Literatura dotycząca intonacji, zorientowana lingwistycznie oraz fonetyczno-akustycznie, podaje zapisy zmian wysokości tonu w mowie, które są bądź wzajemnie nieporównywalne, bądź wręcz sprzeczne. Rozróżnić można zapisy subiektywne i obiektywne. Brak jest dotychczas określenia relacji między tymi dwoma podstawowymi rodzajami zapisu. Metody obiektywne zostały w literaturze przedmiotu sklasyfikowane. Natomiast metody subiektywne, bardzo zróżnicowane, podaje się w ogólnej klasyfikacji w niniejszej pracy. Frazy intonacyjne mogą się różnić pod wpływem wielu współdziałających czynników. Wybrano tutaj dwa źródła zmienności, wpływające na przebieg parametru F_0 , mianowicie (a) długość frazy oraz (b) binarna, segmentalna cecha dźwięczności/bezdźwięczności. Na materiale składającym się z dwóch serii fraz, z systematycznymi zmianami obu czynników, wymówionymi tak przez osoby wykształcone fonetycznie, jak i pozbawione takiego wykształcenia, wykazano, że przy zastosowaniu jednej z metod programowania dynamicznego można usunąć zmienność wynikającą z działania tak jednego, jak i drugiego czynnika. Umożliwia to wykrycie, w danym zakresie zmienności, inwariantów przebiegu F_0 , umożliwiających identyfikację obiektywnych zapisów tego parametru, jako reprezentujących założone ich klasy. Uzyskane wyniki mają podstawowe znaczenie z jednej strony dla syntezy mowy polskiej z reguł, a z drugiej dla automatycznego rozpoznawania i klasyfikacji przebiegów parametru F_0 .

1. Zapis wysokości tonu w mowie.

Istnieją bardzo liczne sposoby zapisu intonacji - zwanej również melodią mowy - oraz jej fizycznego odpowiednika, którym jest zmienna w czasie częstotliwość podstawowa akustycznego sygnału mowy. Jako główne kryterium podziału można przyjąć zróżnicowanie między zapisem subiektywnym oraz obiektywnym. Pierwszy jest wynikiem przetworzenia zjawisk akustyczno-fonetycznych przez słuchacza, a drugi pochodzi z pomiaru wielkości F_0 , jako jednego z kilku zmiennych w czasie parametrów częstotliwościowych, które można ekstrahować z sygnału mowy.

W obrębie tego podstawowego podziału można przeprowadzić dalsze rozróżnienia.

"Opracowano dosiownie setki metod i algorytmów określania wysokości tonu", pisał niedawno Hess (1983; VII, tłum. aut.), który w swej pracy dał pełny przegląd techniczno-fonetycznej problematyki wysokości tonu w mowie, przede wszystkim wyczerpujący przegląd obiektywnych metod ekstrakcji i zapisu czasowo-zmiennej wartości F_0 . Zapis ten, tj. - w przyjętym tutaj rozumieniu - zapis obiektywny, zależy jest od następujących, z góry przyjętych warunków: (a) definicji częstotliwości podstawowej, (b) algorytmu ekstrakcji oraz (c) dokładności aparatury analizującej i zapisującej. Ponieważ istnieje poza tym zwięzły opis metod ekstrakcji i zapisu wysokości tonu tego samego autora (Hess, 1982) oraz nieco wcześniejszy, opracowany przez Schroedera (1976), pominiemy tutaj omówienie zasad i odmian zapisów obiektywnych parametru F_0 . Dokonamy natomiast bardzo ogólnego przeglądu zapisów subiektywnych ze względu na to, że (a) brak jest obecnie takiego uogólniającego przeglądu, (b) jednym z głównych celów wieloletnich prac w zakresie badań nad parametrem F_0 , prowadzonych w ZAF, jest ustalenie relacji między zapisem obiektywnym i subiektywnym oraz (c) w obrębie metod subiektywnych zróżnicowania metodologiczne prowadzą do różnic w wynikach analizy daleko głębszych niż w przypadku zapisów obiektywnych.

2. Ogólna klasyfikacja subiektywnych zapisów wysokości tonu.

Subiektywne zapisy wysokości tonu można sklasyfikować według następujących kryteriów:

1. Podstawowa jednostka opisu.
2. Poziom opisu: strukturalny/substrukturalny.
3. Wykorzystanie pamięci krótkotrwałej/długotrwałej.
4. Zastosowanie/brak wspomaganie audytywnego.
5. Określenie wysokości względnej/bezwzględnej.

2.1. Subiektywne zapisy przebiegów intonacji.

Jako podstawową jednostkę subiektywnego zapisu intonacji bardzo znaczna większość autorów przyjmuje sylabę. Zasada ta jest stosowana, bez doświadczalnego ani teoretycznego uzasadnienia, od ok. r. 1910 do czasów obecnych. Dla języka francuskiego przyjmują ją na przykład Klinghardt i Olbrich (1925 [1911]); Coustenoble i Armstrong (1934) oraz Armstrong (1944), Ward (1948), dla angielskiego Jones (1956), Kingdon (1959), Crystal (1969), O'Connor i Arnold (1973) i Cruttenden (1986); dla włoskiego Chapallaz (1964) i Lichen (1969); dla niemieckiego Trim (1964) i von Essen (1966); dla rosyjskiego Boyanus (1935, 1944) i Ward (1964); dla polskiego Jassem (1962) i Steffen-Batóg (1966).

Poszczególne sylaby równe (tj. takie, w obrębie których wysokość tonu percepcyjnie się nie zmienia) oznacza się kropką albo poziomą kreską, rosnące i opadające - odpowiednio skierowanym odcinkiem prostym, łukiem albo kropką z odpowiednio skierowanym 'ogonkiem', zaś sylaby ze zwrotem kierunku zmiany - dwukrotnie wygiętą kreską albo kropką z podwójnie wygiętym 'ogonkiem' (np. \smile lub \frown). Bez dowodu doświadczalnego prawie wszyscy autorzy przyjmują, że w obrębie jednej sylaby może wystąpić najwyżej jeden zwrot kierunku (intonacja rosnąco-opadająca albo opadająco-rosnąca), a jeśli dwa - to tylko w kolejności: wzrost + spadek + wzrost. Znaki dla poszczególnych sylab umieszcza się najczęściej pomiędzy dwiema liniami poziomymi symbolizującymi odpowiednio górną i dolną granicę normalnego zakresu wysokości tonu w mowie.

Czasem dodaje się trzecią linię pośrednią, odnoszącą się do średniej wysokości (np. Jones 1956 b), a wyjątkowo (Klinghardt i Olbrich 1925) lokuje się intonacyjne znaki sylabiczne względem jednej linii określającej średnio-niski poziom. Wzajemne stosunki między sąsiadującymi sylabami pod względem wysokości tonu odzwierciedlają wzajemne położenia odpowiednich kropek (kreszek itd.).

Brak dotychczas opracowania, które konfrontowałoby subiektywny, sylabiczny zapis intonacji z zapisem obiektywnym i podało choćby wstępne uogólnienia określające korelacje między nimi. Zachodzą tu następujące podstawowe różnice:

(a) Zapis obiektywny ma charakter ciągły z przerwami tylko w miejscach, w których nie występuje ton krtaniowy (ewentualnie takich, w których składowa periodyczna jest tak słaba amplitudowo, że nie można jej w danych warunkach pomiarowych wyekstrahować z sygnału). Zapis obiektywny nie wykazuje zatem podziału na sylaby. Co prawda, w miejscach występowania spółgłosek dźwięcznych, szczególnie obstruentów, pojawiają się załamania przebiegu, nie wystarczają one jednak do wyznaczenia podziału zapisu obiektywnego na odcinki sylabiczne, tym bardziej że interwokaliczne spółgłoski płynne (np. [l, m]) takich załamania często nie wywołują.

(b) Tak oś czasu, jak i oś częstotliwości są w zapisie obiektywnym skalarne. Oś czasu jest zawsze liniowa, natomiast oś częstotliwości bywa liniowa, logarytmiczna lub hiperboliczna, zależnie od stosowanej metody analizy.

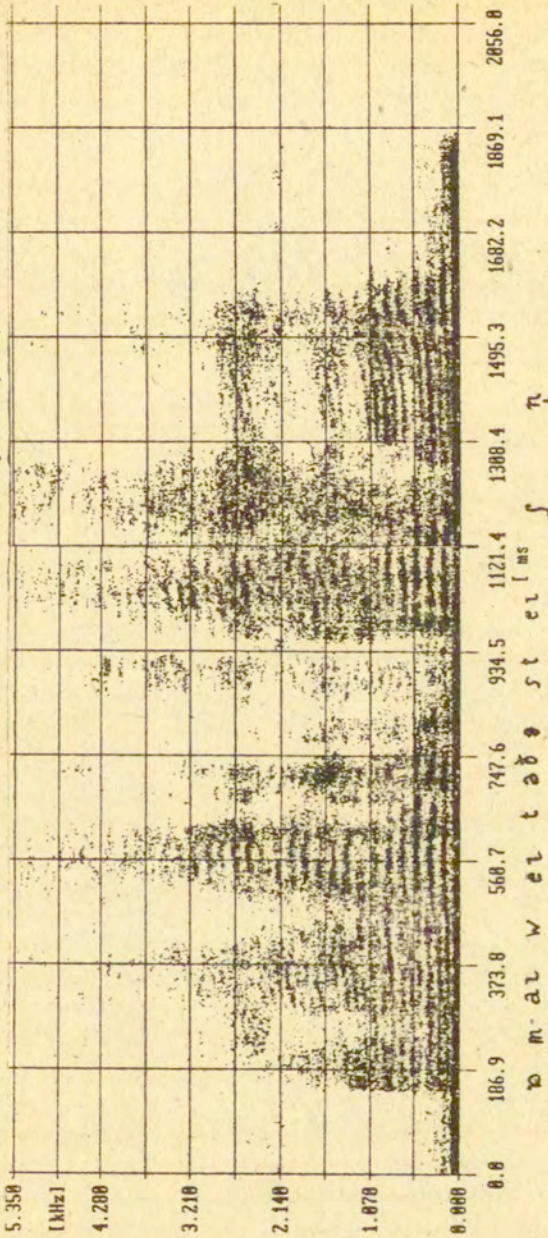
Oś czasu ma zwrot w prawo, jak w piśmie większości języków stosujących pismo alfabetyczne. Jest to oś w zapisie subiektywnym nieskalowana, tj. wyłącznie szeregująca.

Rys. 1a ukazuje skopiowaną z pracy Cruttendena (1986; 40) frazę *On my way to the station...* Rys. 1b przedstawia spektrogram cyfrowy tej samej frazy z zapisu magnetofonowego towarzyszącego publikacji. Widoczna jest niezgodność z zapisem subiektywnym, a mianowicie:

(a) sylaby *to the* są niższe niż sylaba *sta-*

(b) intonacja sylaby *-tion* jest znacznie wyższa niż w zapisie subiektywnym i dochodzi do poziomu sylaby *sta-*

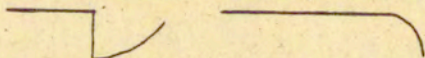
Rys. 1a. Fraza "On my way to the station..." /Gruttenden, 1986; 40/.



Rys. 1b. Spektrogram wypowiedzi "On my way to the station..."

Niektórzy autorzy zapisują subiektywnie przebieg zmian wysokości tonu w mowie za pomocą odcinków linii zakrzywionych odpowiednio w miejscach wzrostu lub spadku, przy czym każdy odcinek odpowiada (nie zdefiniowanej zresztą precyzyjnie) frazie intonacyjnej, np. dla angielskiego Jones (1940), Hill (1965), dla francuskiego Kenning (1983). Tu z kolei niezgodność z zapisem obiektywnym polega przede wszystkim na nieuwzględnieniu przerw wynikających z braku kwaziperiodycznego pobudzenia krtaniowego w miejscach występowania spółgłosek bezdźwięcznych, jak na rys. 2.

Are you coming? Time's getting on.



Rys. 2. Zapis intonacji u Hilla (1965; 49).

Unikalny sposób oznaczania intonacji stosuje we wszystkich swoich pracach D. Bolinger (np. Bolinger 1985), który czyni to w samym składzie literowym, jak to ukazuje przykład na rys. 3.

She ^{nev} said she would er be ^{hi} happy with it.

Rys. 3. Przykład oznaczania intonacji stosowanego przez Bolingera (1985; 17).

Rzadko stosowany jest zapis subiektywny oparty na jednostkach pośrednich między sylabą a frazą intonacyjną, które można określić jako zestroje akcentowe. Taki zapis występuje

m.in. w pracach Pike'a (1945) i Jassema (1952; tu obok systemu sylabowego) dla intonacji angielskiej oraz u Kenning (1963) dla języka francuskiego. Relacje do zapisu obiektywnego są tu analogiczne jak w systemie frazowym. Postacią zapisu są odcinki linii prostej (poziome lub ukośne) albo linie łamane, odpowiednio do kierunku zmian wysokości tonu.

2.2. Klasyfikacja subiektywna przebiegów intonacji.

Na określone pytanie rozpoczynające się od wyrazu *ile* (np. *Ile tam było osób?*) można odpowiedzieć używając różnych liczebników, np. *dwie; pięć; jedna; dziesięć; dwanaście; kilkaset*, w okolicznościach, które nie wymagają zróżnicowania intonacyjnego, czyli w sformułowaniu potocznym 'tym samym tonem'. Mimo to, zapisy subiektywne co najmniej niektórych z tych odpowiedzi będą różne, choćby z uwagi na różną liczbę sylab i lokalizację akcentu. Zapisy obiektywne będą jeszcze bardziej zróżnicowane, ponieważ:

(1) Wyrazy *dwie* i *pięć* są wprawdzie oba jednosylabowe ale pierwszy składa się z samych segmentów dźwięcznych, a drugi rozpoczyna się i kończy segmentami bezdźwięcznymi, zatem - choć oba przebiegi F_0 będą stanowić jedną całość - pierwszy może być dłuższy niż drugi.

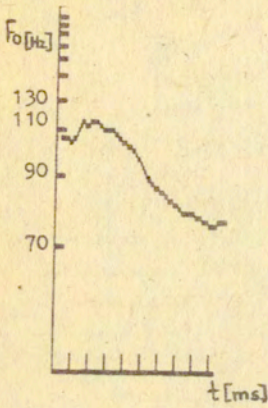
(2) Przebieg F_0 w wyrazie *jedna* będzie wprawdzie (zakładając poprawną ekstrakcję w odrębie zwartej [d]) stanowić jedną całość, ale będzie znacznie dłuższy od przebiegu w wyrazie *dwie*, jako dwusylabowy.

(3) Wyraz *dziesięć*, dwusylabowy, podobnie jak *jedna*, będzie się jednak składać z dwóch kolejnych przebiegów, albowiem wewnątrz występuje segment bezdźwięczny.

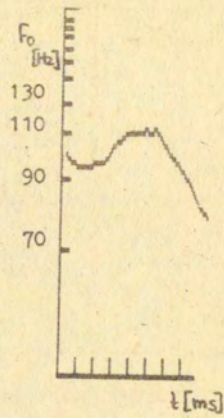
(4) Przebiegi F_0 w wyrazach *dwanaście* i *kilkaset*, choć oba są trzysylabowe, będą różne z uwagi na różne rozkłady segmentów bezdźwięcznych, ale także z powodu różnej lokalizacji akcentu.

Rys. 4 ukazuje omówione zróżnicowania.

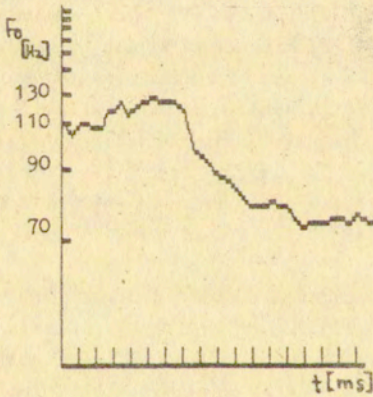
Przykład na rys. 4 ilustruje intuicyjnie zasadne przypuszczenie, że różne przebiegi wysokości tonu w wypowiedziach mogą stanowić warianty realizacyjne jednej klasy przebiegów (czyli jednego elementu strukturalnego), pomiędzy którymi



dwie

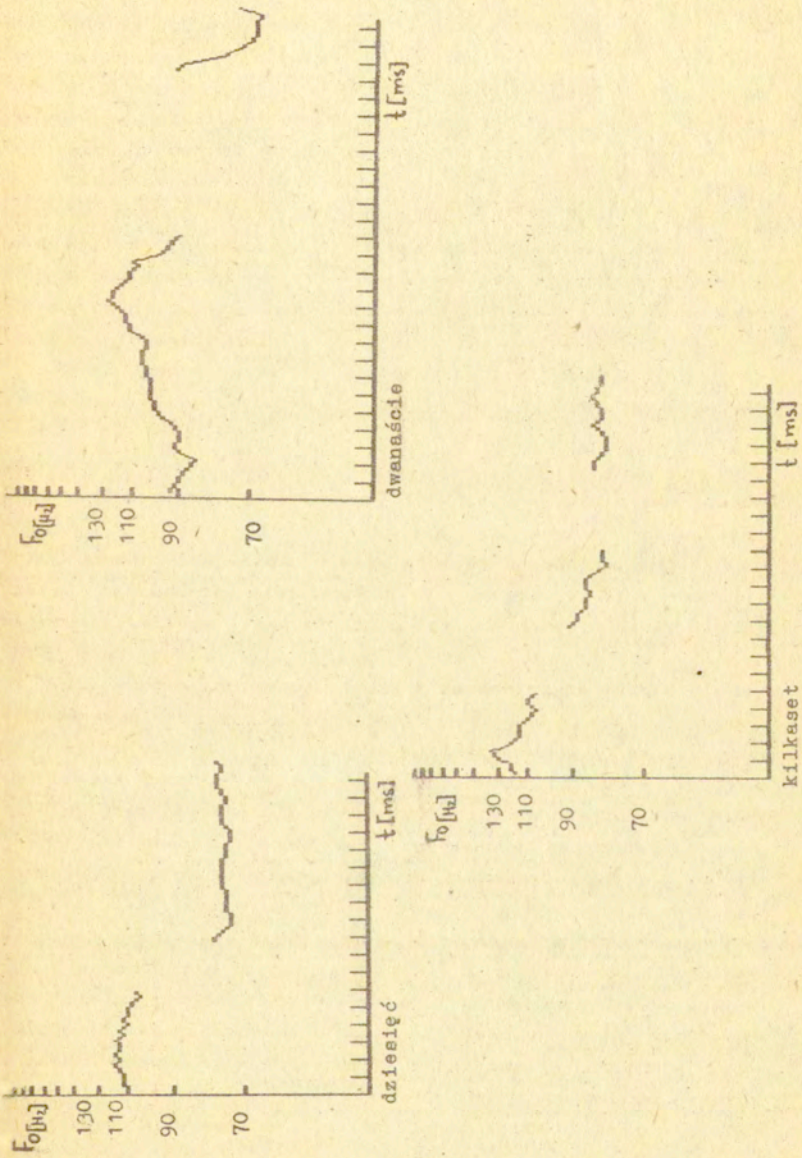


pięć



jedna

Rys.4. Przebiegi parametru F_0 w wypowiedziach "dwie", "pięć", "jedna". Znacznik na skali czasu rozmieszczono co 30 ms.



Rys. 4. Przebiegi parametru F_0 w wypowiedziach "dziesięć", "dwanaście", "kilkaset".
Znacznik na skali czasu rozmieszczono co 30 ms.

zróznicowania są uwarunkowane dającymi się ściśle określić źródłami zmienności substrukturalnej, mianowicie:

- (a) obecnością lub brakiem składowej kwaziperiodycznej, zależnie od substratu segmentalnego wypowiedzi,
- (b) długością wypowiedzi, zależną głównie od liczby sylab oraz
- (c) lokalizacją akcentu.

Jeśli podany powyżej przykład zostanie zrealizowany w eksperymencie, w którym respondentami będą osoby różnej płci i wieku, to zróznicowania pomiędzy poszczególnymi zapisami parametru F_0 będą dodatkowo uzależnione od średniej wysokości głosu oraz jego rozstępu (zakresu zmienności), którymi to cechami będą napewno różnić się między sobą poszczególne głosy, zwłaszcza w zależności od wieku (przy czym w przypadku osób płci męskiej zróznicowania wynikać będą w znacznym stopniu od przejścia mutacji) oraz płci (szczególnie w przypadku osób dorosłych).

Powstaje zatem problem ujęcia w klasy strukturalne różnych przebiegów wysokości tonu, pomiędzy którymi zróznicowania są zdeterminowane i zależą od uwarunkowań kontekstowych, szczególnie w zakresie liczby sylab, obecności składowej periodycznej związanej z substratem segmentalnym wypowiedzenia oraz miejsca akcentu. Te uwarunkowania mają charakter lingwistyczny. Natomiast zmienność zależna od wysokości i zakresu głosu stanowi uwarunkowanie ekstralingwistyczne.

Pierwszą poważniejszą próbę klasyfikacji podjął H. Palmer (1922) i chociaż analogiczne spostrzeżenia o charakterze ogólnikowym i sporadycznym można spotkać w odniesieniu do innych języków, angielski pozostaje tym, w zakresie którego najwięcej w tym względzie uczyniono, przy czym opisy dotyczą przebiegów określonych subiektywnie. Poszczególne opracowania różnią się między sobą pod względem propozycji rozczłonkowania frazy intonacyjnej (lub wypowiedzenia), ale znaczna większość autorów przyjmuje istnienie części konstytutywnej, którą określa się w terminologii angielskiej jako 'nuclear tune' lub 'nuclear tone', czyli 'intonację rdzenną'. W systematycznym opisie grupowania substrukturalnie różnych prze-

biegów w jedną intonację rdzenną bierze się pod uwagę zmienność w zakresie długości przebiegu, określonej liczbą sylab, oraz w zakresie lokalizacji akcentu. Ponieważ spółgłoski bezdźwięczne przypadają w języku angielskim wyłącznie na granicy sylaby, w subiektywnym zapisie sylabicznym z reguły tego źródła zmienności nie bierze się pod uwagę (por. np. Kingdon (1958), O'Coonor & Arnold 1973), ale bywa ono uwzględniane przez niektórych autorów (np. Gimson 1963; 245 ff.).

Metody klasyfikacji obiektów oparte na modelach matematycznych - deterministycznych lub probabilistycznych - można najogólniej podzielić na aprioryczne i aposterioryczne. W pierwszym przypadku stawia się i testuje (na ogół poprzez falsyfikację) z góry postawioną hipotezę dotyczącą liczby klas, do których poszczególne obiekty należą (zazwyczaj są to klasy rozłączne), w drugim zaś przypadku (na ogół poprzez stosowanie metod iteracyjnych) dochodzi się indukcyjnie do liczby (też na ogół rozłącznych) klas. Pierwszego rodzaju metody są mniej obliczeniowo obciążające. Wybór rodzaju metody zależy tak od preferencji badacza, jak i natury problemu. Poniżej zastosowana zostanie jedna z metod apriorycznych, jako że cechy badanego materiału w tym kierunku wydają się jego analizie predestynować.

Jest rzeczą oczywistą, że należy w każdym przypadku określić kryteria, według których ustala się liczebność klas. W przypadku badań nad intonacją kryteria takie nie zostały do obecnej chwili dostatecznie precyzyjnie określone. Skutek tego braku objawia się w bardzo nieraz różnych klasyfikacjach tego samego typu obiektów. I tak dla intonacji brytyjskiej angielszczyzny, dla której istnieje najbardziej obszerna literatura, liczba i rodzaje postulowanych przez poszczególnych autorów intonacji rdzennych jest wyraźnie zróżnicowana, o czym świadczy poniższa, reprezentatywna, ale niekoniecznie wyczerpująca Tablica 1. Zgodnie z dość często stosowaną symboliką oznaczamy intonacje opadające przez F (falling), rosnące przez R (rising), równe przez L (level), rosnąco-opadające przez RF (rising-falling) oraz opadająco-rosnące przez FR (falling-rising).

Tablica 1.

Liczebność klas intonacji rdzennych w języku angielskim
(standardowa odmiana brytyjska) według autorów.

Autor	F	R	L	RF	FR	razem
Palmer (1922)	2	2	0	1	1	6
Jassem (1952)	3	3	2	2	2	12
Kingdon (1958)	4	4	0	4	4	16 *
Faure (1962)	4	4	0	1	1	10 ***
Halliday (1970)	3	2	0	2	2	9
O'Connor & Arnold (1973)	2	2	1	1	1	7 ****
Cruttenden (1986)	2	2	1	1	1	7
Jassem (1987)	3	3	1	2	3	12

* Plus 4 intonacje kombinowane.

** Dopuszcza się możliwość istnienia dalszych jednostek.

*** Plus 1 intonacja kombinowana.

Zróżnicowania w obrębie każdego typu określa się przez relację do dolnej i/lub górnej granicy normalnego zasięgu głosu w mowie albo do rozstępu między początkiem a końcem intonacji rdzennej (np. niska/wysoka intonacja rosnąca, pełna/wysoka/niska intonacja opadająca itp.)

W subiektywnym zapisie strukturalnym stosuje się znaki akcentowo-tonalne dla zaznaczenia intonacji w tekście, przy czym tekst (w płaszczyźnie segmentalnej) zapisywany jest ortograficznie lub fonetycznie. Najczęściej stosowane są znaki syltemu IPA (International Phonetic Association), np.:

(jęz. ang.) 'How 'are you, Mr. ,Richards? (Kingdon, 1958;111)

'Don't I ^know it! (O'Connor & Arnold 1973:229)

(jęz. niem) di: 'klu:ge ,maus (Trim 1964:381)

(jęz. ros.) spa'kojnej 'nofti (Ward 1963:252)

(jęz. pol.) Pójdziemy do ,kina?

Na zło'dzieju ,czapka gore (Jassem 1962;252)

W zapisie strukturalnym opartym na założeniu 4 względnych poziomów wysokości tonu intonacja oznaczona jest, jak w następujących przykładach:

The man did it.

3- °2- -4

Whom did you tell it to yesterday?

°2-3/3- °2- -3/°2- -4 (Pike 1972:75).

2.3. Zapisy subiektywne: pamięć krótko- i długotrwała.

Znaczna większość lingwistycznych opisów intonacji odzwierciedla przeniesiony na modalność optyczną pamięciowy obraz audytywny, jest zatem produktem pamięci długotrwałej. Dotyczy to zresztą również prawie wszystkich opisów segmentalnych.

Dotychczas nie opracowano naukowej metody weryfikacji zapisów pamięciowych fonetycznej formy języka. W zakresie problematyki intonacyjnej nie ma poza tym nawet choćby zbliżonych poglądów co do kryteriów klasyfikacyjnych. Wynika stąd zilustrowana powyżej niezgodność co do liczby i rodzajów klas struktur intonacyjnych.

Sporadycznie spotyka się zapisy subiektywne wykonane przy zastosowaniu pamięci krótkotrwałej. Polegają one na przetworzeniu na modalność wzrokową (optyczną) wrażeń słuchowych uzyskanych przez wielokrotny odsłuch zapisów sygnału mowy (dawniej na płycie gramofonowej, później na taśmie magnetofonowej) w krótkich, jedno- lub kilkusylabowych fragmentach. Prawdopodobnie najwcześniejszym zapisem intonacji, opartym na takiej procedurze, jest praca D. Jonesa (1911). Później stosowali takiego typu słuchową analizę m. in. Jassem (1952; 1963) oraz Crystal i Davy (1969).

Jak wspomniano wyżej, jedną z zasadniczych różnic między zapisem subiektywnym a obiektywnym jest to, że pierwszy operuje tylko względną (i w pewnym zakresie rangową) skalą wysokości, zaś drugi - skalą metryczną. Zasadniczą przyczyną tej różnicy jest brak słuchu bezwzględnego u bardzo znacznej większości słuchaczy. Jednak w kilku pracach zastosowano swego rodzaju kompromis, polegający na wykorzystaniu w pamięci krótkotrwałej pomocy ze strony urządzenia wytwarzającego tony o różnej wysokości, które włącza się bezpośrednio po usłyszeniu fragmentu badanego sygnału mowy i zrównuje się wysokość dostrajając przyrząd do usłyszanej wysokości. Częstotliwość odczytuje się następnie ze skali przyrządu. Dawniej stosowano w takich doświadczeniach widełki stroikowe, a następnie oscylatory elektryczne ze wzmacniaczem. Tego typu metodę 'kwazi-pomiarową' stosowali m. in. D. Jones (1909), Jassem (1952) oraz Cowan (1962).

3. Materiał doświadczalny.

W pracach Jassemę, Demenko & Krzyński (w druku), Surmano-wicz-Demenko (1987) oraz Jassemę (1987) zaproponowano matematyczne metody klasyfikacji przebiegów F_0 w bardzo prostych i krótkich wypowiedzeniach z jednakowym substratem segmentalnym, w całości dźwięcznym, o zbliżonych długościach trwania. Systematycznymi źródłami zmienności były w tamtych pracach: (a) domniemane opozycje lingwistyczne oraz (b) cechy osobnicze w zakresie częstotliwości podstawowej.

Obecna praca rozszerza materiał doświadczalny wprowadzając dwa dalsze źródła zmienności:

(a) długość wypowiedzenia zróżnicowaną poprzez liczbę sylab,

(b) rozczłonkowanie przebiegu wynikające z obecności w substracie segmentalnym fragmentów bezdźwięcznych, tj. PO^- zbalansowanych składowej kwazi-periodycznej.

Materiał doświadczalny składał się z następujących dwóch zestawów wypowiedzi z akcentem na pierwszej sylabie:

(A)

1. Dał (jedna sylaba).
2. Dawał (dwie sylaby).
3. Dawałby (trzy sylaby).
4. Dawałby mi (cztery sylaby).
5. Dawałby mi go (pięć sylab).

(B)

1. Masz (jedna sylaba).
2. Masz to (dwie sylaby).
3. Masz to tu (trzy sylaby).
4. Masz to tutaj (cztery sylaby).

W zestawie (A) każda z wypowiedzi stanowi jeden przebieg parametru F_0 , jako że wszystkie składają się wyłącznie z segmentów dźwięcznych. W zestawie (B) pierwsza wypowiedź kończy się spółgłoską bezdźwięczną, a w następnych spółgłoski bezdźwięczne występują na wszystkich granicach międzysylabicznych.

Powyższe wypowiedzenia zostały wymówione przez głos wzorcowy, każda w czterech intonacjach:

opadającej pełnej - symbol [\],

opadającej niskiej - symbol [,],

rosnącej pełnej - symbol [/],

rosnącej wysokiej - symbol ['].

Sześć osób (w tym 2' przygotowane fonetycznie) powtarzało poszczególne wypowiedzi wzorcowe.

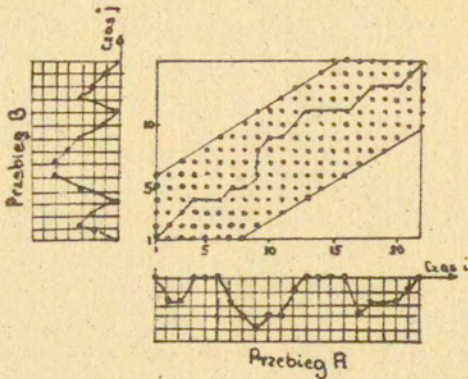
4. Podstawy teoretyczne nieliniowej normalizacji czasowej.

Nierównomierność tempa wypowiedzi jest jedną z przyczyn trudności w rozpoznawaniu sygnału mowy. Z licznych prób rozwiązania tego zagadnienia na uwagę zasługuje najpopularniejsza obecnie metoda "nieliniowego dopasowania czasowego" (time warping), oparta na technice programowania dynamicznego. Aktualnie coraz więcej uwagi przywiązuje się do wykorzystania tej techniki w rozpoznawaniu mowy. Na rys. 5 zilustrowano koncepcję metody nieliniowej normalizacji czasowej w postaci klasycznej (według Sickerta 1983). Nad osiami i oraz j umieszczono dwa przebiegi sygnału A i B reprezentowane przez kolejne próbki mowy

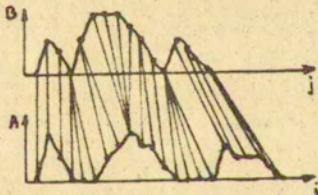
$$A = a(1), \dots, a(i), \dots, a(I);$$

$$B = b(1), \dots, b(j), \dots, b(J).$$

Liczba danych I oraz J określa rozmiary macierzy odległości między poszczególnymi punktami przebiegu A i B. Macierz odległości wynika zwykle z charakteru analizy przebiegów i można ją traktować jako narzuconą. Optymalne dopasowanie obu ciągów nastąpi wówczas, gdy znaleziona zostanie taka ścieżka łącząca lewy dolny i prawy górny róg macierzy, na której suma odległości między elementami $a_{(i)}$ i $b_{(j)}$ będzie minimalna. Celem metody jest więc znalezienie takiego odwzorowania, które optymalnie wyeliminuje różnice czasowe w przebiegach A i B. Problem ten rozwiązano wykorzystując klasyczną technikę optymalizacji, tj. programowanie dynamiczne. Od punktu początkowego (lub końcowego) wylicza się dla wszystkich kolejnych punktów skumulowane optymalne (minimalne) odległości między tymi punktami a punktem końcowym (lub początkowym) według rekursywnej reguły



Rys. 5. Ilustracja nieliniowego czasowego dopasowania przebiegu A oraz B.



Rys. 6. Rezultat nieliniowego czasowego dopasowania przebiegu A oraz B.

$$D_{(i,j)} = d_{(i,j)} + \min D_{(i+1,j)}, D_{(i+1)}, D_{(i,j+1)}$$

gdzie $d_{(i,j)}$ - różnica między segmentem i -tym jednego słowa i segmentem j -tym drugiego.

$D_{(i,j)}$ - minimalna suma odległości między punktem (i,j) a końcowym, reprezentująca optymalną decyzję w punkcie i,j .

Dla podjęcia optymalnej decyzji w punkcie i oraz j wymagana jest znajomość decyzji podjętych w trzech bezpośrednio poprzedzających punktach. Na rys. 6 przedstawiono rezultat nieliniowego dopasowania czasowego przebiegów A i B.

Omawiana procedura ma wiele odmian, przy czym zasadnicze ulepszenia mają na celu głównie zmniejszenie pracochłonności obliczeń. Zamiast wyliczać wszystkie odległości $D_{(i,j)}$ można analizować ograniczone pole macierzy odległości. O łącznym czasie obliczeń decyduje:

- liczba analizowanych punktów na siatce współrzędnych,
- przyjęta miara odległości i sposób jej obliczania,
- postać równania programowania dynamicznego.

5. Realizacja techniczna metody normalizacji nieliniowej przebiegów częstotliwości podstawowej.

5.1. Struktura oprogramowania.

Wykorzystano następujące programy:

1. Program "nag", umożliwiający wczytanie sygnału do komputera (opracowany uprzednio w Pracowni).
2. Program "pomiar" mający na celu wprowadzenie wyników pomiaru parametru F_0 z tonometru do pamięci komputera i zapamiętanie ich w plikach danych.
3. Program "anfo" umożliwiający opracowanie wyników pomiaru częstotliwości podstawowej i wyświetlenie przebiegów parametru F_0 na ekranie monitora.
4. Program "opty" realizujący algorytm programowania dynamicznego.

Program "pomiar" zawiera procedurę obliczania długości okresu sygnału prostokątnego otrzymanego z tonometru i wyprowadzania wyniku w formie wykresu na ekran monitora lub drukarkę. Program przetestowano przy użyciu generatora oraz miernika

częstotliwości. Częstotliwość próbkowania sygnału wynosi 10.7 KHz. Wyniki pomiaru zaokrąglono do 1 Hz. Dokładność i poprawność pomiaru uzależniona jest od pracy tonometru. Program "anfo" umożliwia analizę wyników pomiaru częstotliwości podstawowej. W przypadku pojawienia się dużego błędu pomiaru dokonywana jest korekcja, polegająca na zastąpieniu wyniku błędnego wartością średnią z sąsiednich wartości długości okresu. Opcjonalnie program z wiera procedury: (1) wygładzania przebiegu (okno Hanna) według zależności

$$w_q = 0.25 * x(q-1) + 0.5 * x(q) + 0.25 * x(q+1),$$

gdzie x_q - próbki wygładzanego przebiegu

(2) obliczania wartości średnich parametru F_0 w różnych przedziałach czasowych oraz (3) wyświetlania przebiegów na ekranie monitora.

Program normalizacji nieliniowej "opty" zrealizowano na podstawie algorytmu programowania dynamicznego (przedstawionego w p. 4). Z uwagi na charakter analizowanego parametru F_0 konieczne było wprowadzenie zasad tworzenia macierzy odległości. Przyjęto, że odległość między przebiegami i oraz j będzie określać tablica różnic logarytmów długości okresów obliczanych zgodnie z zależnością:

$$d_{ij} = \ln(t_i) - (\ln(t_j + q))$$

gdzie $i = 1 \dots I$

I - liczba okresów i -tego przebiegu

t_i - czas trwania i -tego okresu

$j = 1 \dots J$

J - liczba okresów j -tego przebiegu

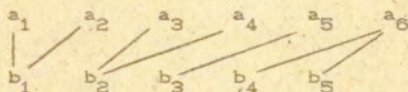
t_j - czas trwania j -tego okresu.

Zastosowanie skali logarymicznej umożliwia analizę zmian względnych, nie eliminuje jednak różnic wynikających z odmiennych zakresów zmienności parametru. Poprzez przesuwanie jednego z przebiegów wzdłuż skali częstotliwości (dodawanie lub odejmowanie pewnej stałej wartości q do długości poszczegól-

gólnych okresów) można znaleźć takie położenie obu przebiegów względem siebie, przy którym odległość między nimi będzie minimalna. W skali liniowej operacja ta będzie równoważna pomnożeniu lub podzieleniu danych przebiegu przez wartość stałą.

5.2. Przykłady normalizacji nieliniowej.

Przykładowo poddano normalizacji ciąg A zawierający 6 danych: 100, 200, 300, 400, 500, 600 oraz ciąg B składający się z 5 danych: 100, 400, 500, 600. W tabelicy 2 przedstawiono odległości między poszczególnymi elementami ciągu A i B obliczone, zgodnie z zależnością (3), przy wartości $q = 0$. W tabelicy 3 przedstawiono macierz odległości optymalnych. Otrzymano następujące przyporządkowanie elementów:



Rys. 7. Przykład przyporządkowania elementów ciągu A i B.

Odległość między przebiegami określa wartość $G_0 = 1.13$. W tabelicach 4 i 5 przedstawiono rezultaty normalizacji przy wartości $q = -1$ a w tabelicach 6 i 7 przy wartości $q = -0.22$. Zauważyć można, że optymalny wynik normalizacji osiągnięto przy wartości $q = -0.22$ (wartość G_0 jest minimalna i równa 0.87). Na rys. 8 przedstawiono rezultat normalizacji przebiegów częstotliwości podstawowej w naturalnych wypowiedziach w gipsie męskim i kobiecym. W tabelicy 8 pokazano wartości parametru F_0 w badanych przebiegach, a w tabelicy 9 wyniki normalizacji. Przebiegi znormalizowano względem wartości minimalnej. Dla zmniejszenia różnic między przebiegami należy zmienić położenie przebiegów względem siebie (tzn. dodać stałą q do wartości jednego z przebiegów). Odległość między przebiegami częstotliwości podstawowej określają więc dwie dane: wartość G_0 wynikająca z różnic w charakterze zmienności parametru F_0 oraz wartość q - składowa stała wynikająca z różnic w zakresach skal. Na rys. 9 przedstawiono ogólną strukturę programu "opty".

Tablica 2. Macierz odległości między ciągiem A oraz B

		ciąg A					
		a_1	a_2	a_3	a_4	a_5	a_6
		100	200	300	400	500	600
b_1	100	0	0,69	1,10	1,39	1,61	1,79
b_2	400	1,39	0,69	0,29	0	0,22	0,41
b_3	500	1,61	0,92	0,51	0,22	0	0,18
b_4	600	1,79	1,10	0,69	0,41	0,18	0
b_5	700	1,95	1,25	0,85	0,56	0,34	0,15

Tablica 3. Macierz odległości skumulowanych.

W lewym górnym rogu macierzy podkreślono wartość G_0 .

		a_1	a_2	a_3	a_4	a_5	a_6
b_1	<u>1,13</u>	1,13	1,25	1,54	1,99	2,53	
b_2	2,52	1,13	0,44	0,15	0,38	0,74	
b_3	3,41	1,80	0,89	0,38	0,15	0,34	
b_4	4,33	2,53	1,44	0,74	0,34	0,15	
b_5	5,10	3,15	1,90	1,05	0,49	0,15	

Tablica 4. Macierz odległości między ciągiem A i B / $q=-1/$

1,00	1,69	2,1	2,39	2,61	2,79
0,39	0,31	0,71	1,00	1,22	1,41
0,61	0,008	0,49	0,78	1,00	1,18
0,79	0,1	0,31	0,59	0,82	1,0
0,95	0,25	0,15	0,44	0,66	0,85

Tablica 5. Macierz odległości skumulowanych / $q=-1/$.

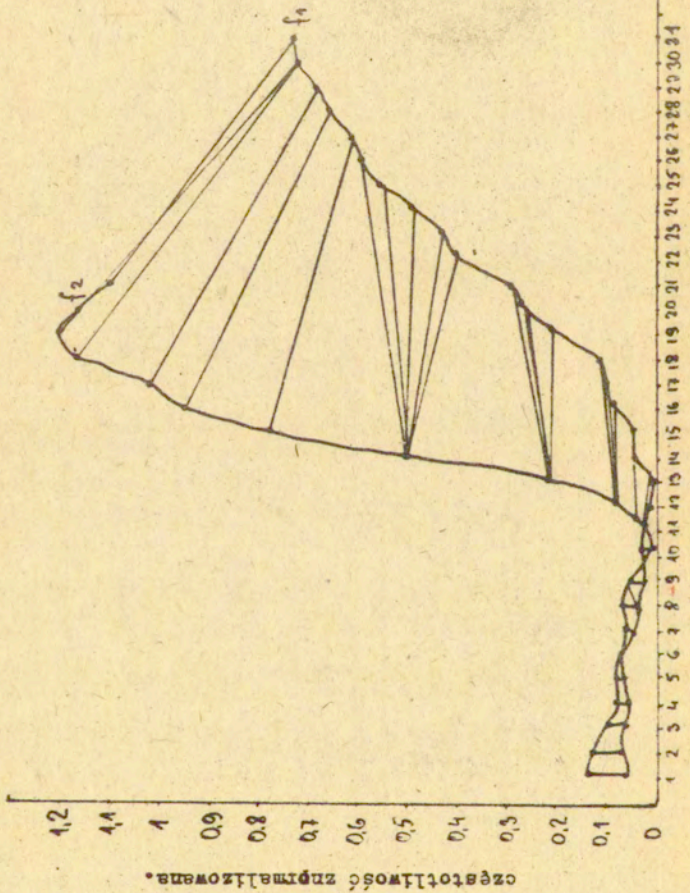
3.59	4.28	5.25	5.83	6.50	7.23
2.67	2.59	3.15	3.44	3.89	4.43
2.81	2.28	2.59	2.44	2.66	3.03
2.99	2.20	2.26	2.10	1.66	1.85
3.30	2.36	2.10	1.95	1.51	0.85

Tablica 6. Macierz odległości między ciągiem A i B / $q=-0,22/$

0,22	0,91	1,32	1,61	1,83	2,01
1,17	0,47	0,07	0,22	0,44	0,63
1,39	0,7	0,29	0	0,22	0,4
1,57	0,88	0,47	0,19	0,04	0,22
1,73	1,03	0,63	0,34	0,12	0,07

Tablica 7. Macierz odległości skumulowanych / $q=-0,22/$.

<u>0,87</u>	1.09	1.49	1.93	2.60	3.33
1.81	0.65	0.17	0.33	0.77	1.31
2.48	1.09	0.40	0.11	0.32	0.69
3.21	1.64	0.76	0.29	0.10	0.29
3.91	2.18	1.15	0.52	0.18	0.07



Rys. 8. Wynik normalizacji czasowej przebiegu parametru F_0 w wypowiedzi "dobrze"
/f₁ - głos męski, f₂ - głos kobiecy/.

Tablica 8.

Wartości częstotliwości podstawowej w przebiegu f_1 oraz f_2
/w Hz/.

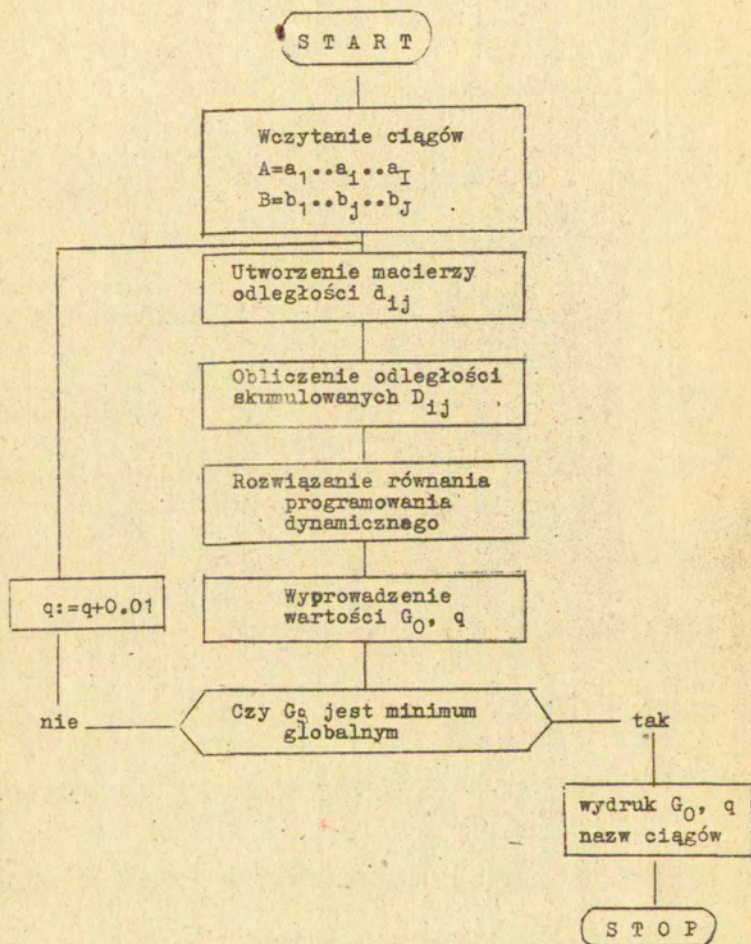
f_1 : 74, 75, 73, 74, 87, 74, 73, 73, 71, 69, 70, 76, 88, 115.
152, 182, 208, 226, 230, 226, 208.

f_2 : 207, 203, 193, 193, 193, 187, 183, 185, 180, 179, 183, 178,
178, 186, 188, 198, 199, 203, 221, 231, 235, 267, 273, 293,
313, 322, 334, 350, 357, 365, 375.

Tablica 9.

Optymalna ścieżka określająca przyporządkowanie elementów
w ciągu f_1 oraz f_2 .

Min=3.418 a=2 b=2
Min=3.370 a=3 b=3
Min=3.345 a=4 b=3
Min=3.334 a=5 b=4
Min=3.183 a=6 b=5
Min=3.172 a=7 b=6
Min=3.165 a=8 b=6
Min=3.158 a=9 b=7
Min=3.157 a=9 b=8
Min=3.147 a=9 b=9
Min=3.130 a=9 b=10
Min=3.107 a=9 b=11
Min=3.106 a=10 b=12
Min=3.106 a=10 b=13
Min=3.106 a=11 b=14
Min=3.076 a=11 b=15
Min=3.036 a=12 b=16
Min=3.026 a=12 b=17
Min=3.011 a=12 b=18
Min=2.977 a=13 b=19
Min=2.950 a=13 b=20
Min=2.932 a=13 b=21
Min=2.898 a=14 b=22
Min=2.792 a=14 b=23
Min=2.709 a=14 b=24
Min=2.697 a=14 b=25
Min=2.643 a=14 b=26
Min=2.561 a=15 b=27
Min=2.401 a=16 b=28
Min=2.107 a=17 b=29
Min=1.700 a=18 b=30
Min=1.258 a=19 b=30
Min=0.800 a=20 b=30
Min=0.358 a=21 b=32



Rys. 9. Ogólna struktura programu "opty".

6. Przykłady.

W tablicach 10 i 11 przedstawiono czasy trwania poszczególnych wypowiedzi wzorcowych dla badanych typów intonacji. Różnice w długościach wypowiedzi wahają się w granicach 4000 - 12300 ms dla wypowiedzi z zestawu A oraz w granicach 3400 - 10600 ms dla wypowiedzi z zestawu B. W tablicy 12 przedstawiono wyniki normalizacji przebiegów parametru F_0 w wypowiedziach wzorcowych z zestawu A. Porównywano ze sobą przebiegi w obrębie tego samego typu intonacji oraz w obrębie 2 różnych typów intonacji odpowiednio rosnących lub opadających. Nie przedstawiono rezultatów porównywania ze sobą przebiegów rosnących oraz opadających, ponieważ różnice między nimi są ewidentne. Odległość G_0 w obrębie tego samego typu intonacji jest znacznie mniejsza niż dla dwóch różnych typów intonacji. Przykładowo, przy porównywaniu przebiegów parametru F_0 w wypowiedziach *dał* (nr.1) i *dawał* (nr.2) reprezentujących intonację pełną opadającą otrzymano wartość $G_0 = 41 \times 10^{-2}$ (przy wartości $q=0.01$ zapisanej w tablicy 12 poniżej wartości G_0), natomiast przy porównywaniu przebiegów parametru F_0 w wypowiedziach *dał* (nr.1) oraz *dawał* (nr.6) reprezentujących intonację pełną i niską opadającą uzyskano wartość $G_0 = 178 \times 10^{-6}$ (przy wartości $q=-0.14$). Można przyjąć, że różnice w długościach wypowiedzi zostały zlikwidowane. Przykładowo, przebiegi parametru F_0 w wypowiedziach *dał* (nr.1) oraz *dawałby mi go* (nr.5) - reprezentujące intonację pełną opadającą są do siebie bardziej podobne (wartość $G_0 = 85 \times 10^{-2}$) niż przebiegi parametru F_0 w wypowiedziach *dał* (nr.1) oraz *dał* (nr.6) reprezentujące intonację opadającą pełną i niską (wartość $G_0 = 178 \times 10^{-2}$). Zauważyć również można wpływ własności segmentalnych wypowiedzi na wynik porównywania przebiegów parametru F_0 (tablica 12). Przebiegi o podobnej długości (strukturze segmentalnej) są z reguły do siebie bardziej podobne niż przebiegi o różnych długościach (różniące się własnościami segmentalnymi). Przykładowo, przebieg nr.1 (*dał*) jest bardziej podobny do przebiegu nr.2 (*dawał*, wartość $G_0 = 41 \times 10^{-2}$) niż do przebiegu nr.5 (*dawałby mi go*, wartość $G_0 = 85 \times 10^{-2}$) i bardziej podobny do przebiegu nr.6 (*dał*, intonacja niska opadająca, wartość $G_0 = 178 \times 10^{-2}$) niż do przebiegu nr.10 (*dawałby mi go*, intonacja niska opadająca, wartość $G_0 = 360 \times 10^{-2}$). Wartości G_0 są wyższe w grupie przebiegów rosnących ponieważ

zakres zmienności parametru był większy dla przebiegów rosnących niż opadających. Przykłady przebiegów parametru F_0 w wypowiedziach wzorcowych z zestawu A przedstawiono na rys. 10 i 11. Podobne porównanie przebiegów parametru F_0 przeprowadzono dla wypowiedzi z zestawu B. Wyniki przedstawiono w tablicy 13. Zauważyć można analogie wyników ukazanych w tablicach 12 i 13 i uogólnić przedstawione wnioski dla wypowiedzi z zestawu A i B.

Rys. 12 i 13 prezentują przebiegi parametru F_0 dla wypowiedzi z zestawu B. W tablicy 14 przedstawiono wynik porównywania przebiegów parametru F_0 w wypowiedziach z zestawu A i zestawu B. Wartość G_0 waha się w granicach 38×10^{-2} - 166×10^{-2} . Zauważyć można, że większą wartość G_0 otrzymano przy porównywaniu przebiegów dłuższych (4 lub 5 sylabowych) niż krótkich (1 lub 2 sylabowych). W tablicy 15 przedstawiono rezultat porównywania przebiegu parametru F_0 w wypowiedziach z zestawu A dla osoby dość poprawnie powtarzającej wypowiedzi wzorcowe (Głos WI). Wyniki w tablicy 15 są podobne do wyników w tablicy 12. Jedynie w grupie przebiegów opadających można zauważyć większą niezgodność. Przebiegi nr. 4 i 5 (intonacja pełna opadająca) są bardziej podobne do przebiegu nr. 10. (intonacja niska opadająca), wartość $G_0 = 122 \times 10^{-2}$ i $G_0 = 114 \times 10^{-2}$, niż do niektórych przebiegów reprezentujących intonację pełną opadającą (np. do przebiegu nr. 1 - wartość $G_0 = 120 \times 10^{-2}$). Wynik ten wydaje się uzasadniony, ponieważ przebiegi nr. 4 i 5 są bardziej podobne wizualnie do przebiegu 10 niż do przebiegu nr. 1 (por. rys. 14). W tablicy 16 przedstawiono wynik porównania przebiegów parametru F_0 w wypowiedziach z zestawu A dla głosu WI oraz w wypowiedziach wzorcowych z zestawu A. W tablicy 17 widnieją rezultaty normalizacji przebiegów parametru F_0 w wypowiedziach osoby, która gorzej imitowała wypowiedzi wzorcowe (głos HK). W tablicy 17 można zauważyć, że osoba ta rozróżniała poprawnie tylko 2 typy intonacji (intonację rosnącą i opadającą por. rys. 15). Wyniki normalizacji intonacji zarówno w wypowiedziach wzorcowych, jak i w wypowiedziach osób nie szkolonych fonetycznie wykazały możliwość eliminacji wpływu na przebiegi parametru F_0 dwóch źródeł zmienności, mianowicie (a) długości frazy oraz (b) segmentalnej cechy dźwięczności/bezdźwięczności.

Tablica 10. Długości wypowiedzi w zestawie A.

Lp.	Wypowiedź	Długość wypowiedzi w/ms			
		Typ intonacji			
		↘	↘	↘	↘
1	dał	4150	4200	4000	4100
2	dawał	5950	6300	5940	5900
3	dawałby	8100	7800	7300	8600
4	dawałby mi	10100	9500	9100	9600
5	dawałby mi go	12300	12000	10550	11300

Tablica 11. Długości wypowiedzi w zestawie B.

Wypowiedź		ma	szt	s	t	u	t	aj
masz		2950						
masz to		1800	2550	2100				
masz to		1850	1950	1200	1600	1900		
tu								
masz to	tutaj	1750	2300	1150	950	1250	1050	2400
masz		2400						
masz to		1900	2500	1900				
masz to	tu	1950	1900	1350	1550	1850		
masz to	tutaj	1850	1900	1900	1000	1050	1350	2250
masz		2900						
masz to		2250	2400	2100				
masz to	tu	2200	1950	1150	1500	2150		
masz to	tutaj	1950	1750	1750	1050	1050	1300	2850
masz		2800						
masz to		1650	2650	2300				
masz to	tu	1950	2100	1450	1450	2100		
masz to	tutaj	1800	1850	1850	1000	900	1350	2800

Typ intonacji		Typ intonacji									
		Opadająca niska					Opadająca pełna				
		Numer przebiegu					Numer przebiegu				
1	2	3	4	5	6	7	8	9	10		
1	x	41 0,01	64 -0,01	72 -0,01	85 -0,01	178 -0,14	180 -0,11	210 -0,12	219 -0,13	360 -0,12	
2	x	x	68 0,03	69 -0,05	70 -0,05	204 -0,21	251 -0,18	253 -0,24	213 -0,15	333 -0,09	
3		x	63 0,02	63 0,02	84 0,08	234 -0,24	212 -0,24	202 -0,25	166 -0,17	270 -0,24	
4				x	70 0,02	330 -0,32	194 -0,1	311 -0,31	203 -0,21	305 -0,26	
5					x	363 -0,31	376 -0,27	311 -0,31	225 -0,18	343 -0,31	
6						x	38 0,04	41 0,01	71 0,07	80 0,05	
7							x	57 -0,01	65 0,05	77 -0,01	
8								x	71 0,09	81 0,01	
9									x	79 -0,06	
10										x	

Tablica 12. Wyniki normalizacji dla wypowiedzi wzorcowych z zestawu A.
/ W tablicy podano wartości σ_0 przemnożone przez 10^{-2} /.

Typ intonacji		Typ intonacji									
		Rosnąca niska					Rosnąca pełna				
Typ intonacji		Rosnąca niska					Rosnąca pełna				
		Numer przebiegu									
		11	12	13	44	15	16	17	18	19	20
11	x	138	-0,09	91	156	184	197	330	325	491	389
12		x		91	127	173	452	529	554	640	-0,29
13				0,01	-0,05	0,01	-0,13	-0,2	-0,11	-0,21	548
14			x	x	54	166	537	552	530	610	-0,12
15				0,09	0,02	0,09	-0,29	-0,37	-0,37	-0,44	571
16				-0,03	x	138	350	440	344	416	-0,41
17						-0,03	-0,29	-0,34	-0,31	-0,34	355
18						x	385	446	372	441	-0,29
19							-0,21	-0,34	-0,21	-0,31	350
20							x	99	153	163	-0,14
								-0,03	0,01	-0,06	134
								x	158	128	0,01
									0,02	0,01	152
									x	101	103
										-0,02	0,02
										x	101
											0,09
											x
											0,09
											x

Tabela 12. Wyniki normalizacji dla wypowiedzi wzorcowych z zestawu A.
/ W tablicy podano wartości G_0 przemnożone przez 10^{-2} /.

Typ intonacji		Typ intonacji							
		Opadająca niska				Opadająca pełna			
Typ intonacji	Opadająca niska	Numer przebiegu							
		1	2	3	4	5	6	7	8
1	x	41 -0,03	93 -0,005	120 -0,02	204 -0,12	159 -0,4	240 -0,06	316 -0,15	
2		x	50 -0,01	139 0,03	260 -0,01	255 -0,01	297 0,06	254 0,02	
3			x	89 0,06	206 -0,6	219 -0,13	213 -0,02	157 0,09	
4				x	301 -0,09	193 -0,09	141 0,08	170 -0,04	
5					x	45 0,003	90 0,005	77 0,002	
6						x	56 0,007	57 0,001	
7							x	41 0,001	
8								x	

Tabela 13. Wyniki normalizacji dla wypowiedzi wzorcowych z zestawu B.

Typ intonacji		Typ intonacji															
		Rosnąca niska								Rosnąca pełna							
Typ intonacji	Rosnąca niska	Numer przebiegu															
		Numer przebiegu								Numer przebiegu							
		9	10	11	12	13	14	15	16	9	10	11	12	13	14	15	16
		x	76 -0,14	94 -0,12	137 -0,1	284 -0,24	346 -0,42	406 -0,44	383 -0,31								
	Rosnąca niska		x	152 0,07	151 0,08	331 -0,006	307 -0,3	357 -0,3	417 -0,4								
				x	132 0,007	190 -0,12	194 -0,11	272 -0,12	318 -0,11								
					x	327 -0,08	393 -0,03	328 -0,01	295 0,25								
						x	80 0,08	80 0,01	99 0,05								
							x	70 -0,05	88 0,08								
	Rosnąca pełna							x	85 0,005								
									x								
										x							

Tabela 13. Wyniki normalizacji dla wypowiedzi wzorcowych z zestawu B.

Tablica 14. Porównanie przebiegów parametru F_0 w wypowiedziach zestawu A oraz B.

Typ intonacji	Wypowiedź	Wartość	
		$G_0 \times 10^{-2}/$	q
Typ intonacji	masz - dał	73	0,15
	masz to - dawał	112	0,03
	masz to tu - dawałby	105	0,025
	masz to tutaj - dawałby mi	151	0,17
	masz - dał	47	0,04
	masz to - dawał	65	0,05
	masz to tu - dawałby	70	0,07
	masz to tutaj - dawałby mi	85	0,05
	masz - dał	56	-0,07
	masz to - dawał	84	-0,05
	masz to tu - dawałby	96	0,02
	masz to tutaj - dawałby mi	166	0,04
Typ intonacji	masz - dał	38	-0,05
	masz to - dawał	69	0,01
	masz to tu - dawałby	94	-0,01
	masz to tutaj - dawałby mi	96	-0,03

Tabela 16. Wyniki porównania przebiegów parametru F_0 w wypowiedziach wzorcowych oraz płosie WI.

LP	Wypowiedź	Wartość G_0	
		G_0	q
Typ intonacji	1 dał	35	-0,01
	2 dawał	72	-0,12
	3 dawałby	72	0,08
	4 dawałby mi	87	-0,17
	5 dawałby mi go	129	0,15
	6 dał	30	-0,03
	7 dawał	35	-0,01
	8 dawałby	49	0,03
	9 dawałby mi	117	-0,05
	10 dawałby mi go	134	0,19
	11 dał	46	0,01
	12 dawał	61	0,15
	13 dawałby	106	0,02
	14 dawałby mi	100	0,18
	15 dawałby mi go	109	-0,05
	16 dał	69	-0,01
	17 dawał	47	0,03
	18 dawałby	104	0,05
	19 dawałby mi	110	-0,07
	20 dawałby mi go	119	0,02

Typ intonacji		Typ intonacji									
		Opadająca niska					Opadająca pełna				
		Numer przebiegu					Numer przebiegu				
		11	12	13	14	15	16	17	18	19	20
Opadająca niska	Numer przebiegu	x	88 0,1	120 0,03	161 -0,04	190 -0,09	390 -0,11	399 -0,07	339 -0,11	447 -0,13	514 0,003
			x	134 0,02	157 -0,09	172 -0,02	500 -0,14	486 -0,12	445 -0,1	527 -0,006	692 -0,007
				x	160 -0,06	150 -0,09	545 -0,14	514 -0,14	401 -0,1	493 -0,14	694 -0,12
					x	92 -0,05	340 -0,11	308 -0,06	250 -0,04	287 -0,07	716 -0,08
						x	334 -0,13	292 -0,16	234 -0,09	302 -0,036	562 -0,01
							x	79 0,03	98 0,01	108 0,06	168 0,07
								x	113 0,06	180 -0,02	185 -0,03
									x	145 0,02	174 -0,05
										x	176 -0,02
											x

Tablica 15. Wyniki normalizacji dla wypowiedzi z zestawu A. Głos WI

Typ intonacji		Typ intonacji									
		Opadająca niska					Opadająca pełna				
		Numer przebiegu					Numer przebiegu				
	1	2	3	4	5	6	7	8	9	10	
Opadająca niska	1	x	62 -0,4	68 0,02	83 -0,03	120 -0,07	137 -0,14	141 -0,12	150 -0,1	161 -0,11	164 -0,1
	2		x	76 0,03	77 -0,001	118 -0,03	215 -0,17	186 -0,003	186 -0,14	293 -0,008	171 0,001
	3			x	86 -0,08	113 -0,08	255 -0,24	217 -0,24	205 -0,19	210 -0,19	168 -0,14
	4				x	66 -0,02	216 -0,22	197 -0,19	161 -0,16	167 -0,17	122 -0,008
	5					x	244 -0,2	205 -0,16	166 -0,12	172 -0,13	114 -0,005
	6						x	42 0,04	65 0,05	66 0,05	134 0,14
	7							x	58 0,02	65 0,11	102 0,1
	8								x	54 0,02	99 0,06
	9									x	80 0,2
	10										x
Opadająca pełna	Numer przebiegu										
Opadająca niska	Numer przebiegu										

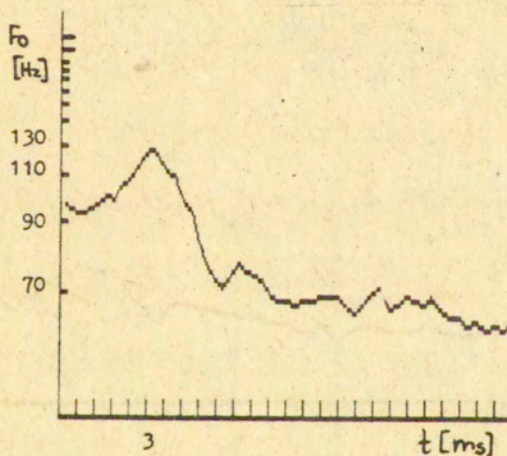
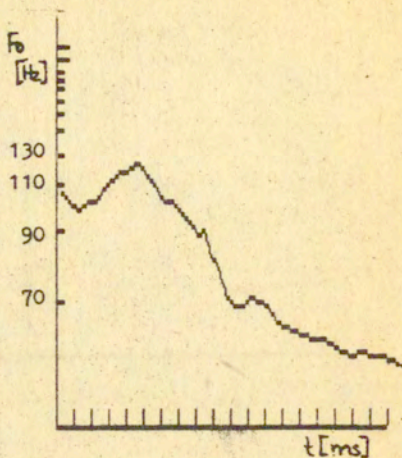
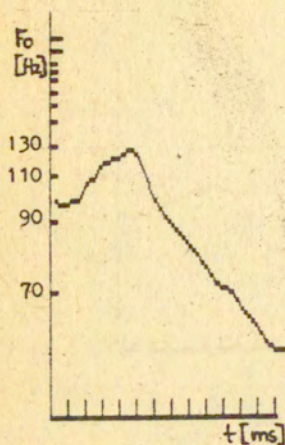
Tablica 15. Wyniki normalizacji dla wypowiedzi z zestawu A, Gros WI.

Typ intonacji		Typ intonacji									
		Opadająca niska					Opadająca pełna				
Typ intonacji	Opadająca niska	Numer przebiegu									
		1	2	3	4	5	6	7	8	9	10
1	x	55 -0,01	72 0,05	118 0,07	111 0,07	69 -0,02	78 -0,03	100 0,02	114 0,01	128 0,01	
2	x	78 0,01	95 0,04	115 -0,01	95 0,02	120 0,02	169 0,03	144 0,02	206 0,09		
3		x	76 -0,02	104 0,01	111 -0,07	116 -0,07	124 -0,1	166 -0,12	172 -0,1		
4			x	93 -0,02	100 -0,09	106 -0,08	137 -0,05	160 -0,14	185 -0,15		
5				x	105 -0,08	116 -0,01	110 -0,09	156 -0,1	125 -0,1		
6					x	40 -0,01	71 -0,07	130 -0,12	123 -0,06		
7						x	73 -0,07	132 -0,13	123 0,05		
8							x	96 -0,04	80 0,35		
9								x	93 0,03		
10									x		

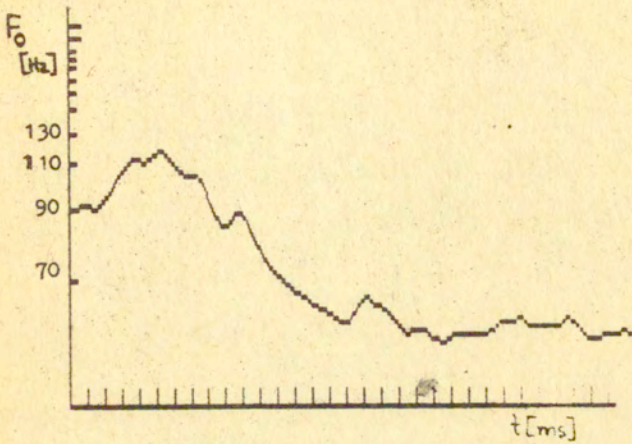
Tablica 17. Wyniki normalizacji dla wypowiedzi z zestawu A. Głos HK.

		Typ intonacji											
		Rosnąca niska					Rosnąca pełna						
		Numer przebiegu											
	11	12	13	14	15	16	17	18	19	20			
Rosnąca niska		x	0,07	0,1	52	90	0,1	187	0,08	111	180	96	193
Numer przebiegu		x		64	0,01	-0,02	0,1	169	208	175	119	103	105
				x		86	0,04	156	0,08	143	115	99	144
						-0,04	0,08	197	0,04	88	76	80	90
						x		0,04	-0,17	0,03	0,01	0,01	0,02
								x	0,01	92	77	135	188
									0,03	0,02	0,08	0,08	0,02
									x	131	162	90	143
										-0,09	0,1	-0,17	0,15
										x	121	76	141
										0,2	-0,02	0,21	0,21
										x	100	84	84
											0,06	0,01	0,01
											x	195	195
												x	0,17
													x

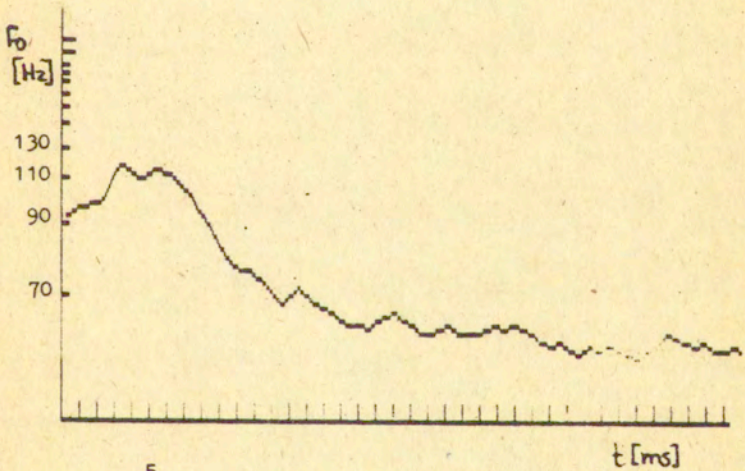
Tabela 17. Wyniki normalizacji dla wypowiedzi z zestawu A.Gros HK.



Rys. 10. Wypowiedzi wzorcowe: 1 - dał, 2 - dawał, 3 - dawałby
Intonacja pełna opadająca. Znacznik na osi czasu
rozmieszczono co 30 ms.

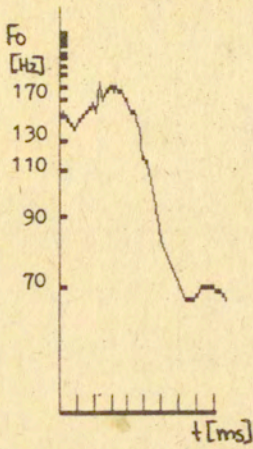


4

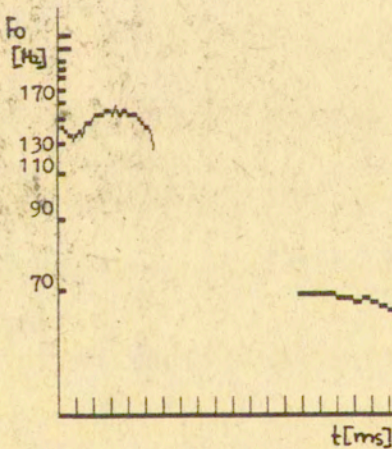


5

Rys. 11. Wypowiedzi wrzycowe: 4 - dawałby mi, 5 - dawałby mi go. Intonacja pełna opadająca. Znacznik na skali czasu rozmieszczono co 30 ms.

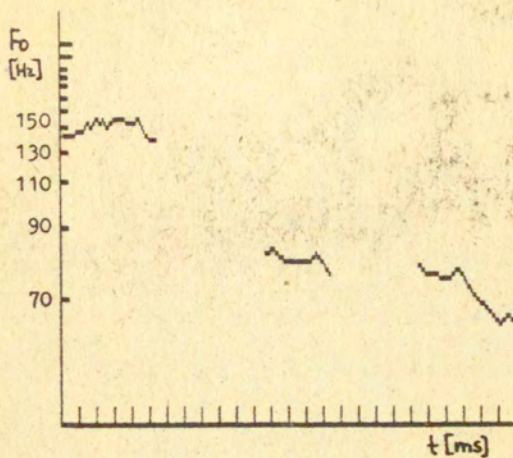


1

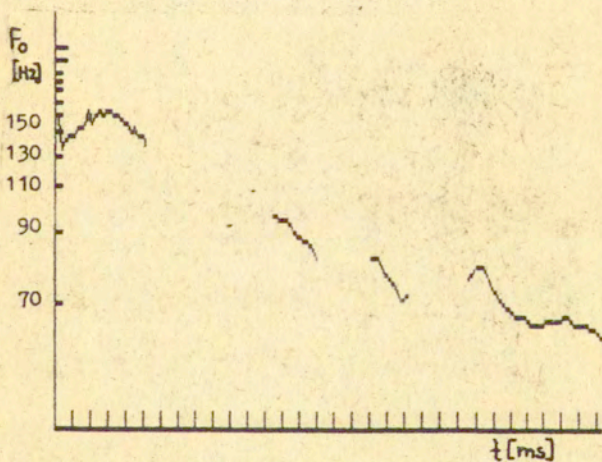


2

Rys. 12. "wypowiedzi wzorcowe: 1 - masz, 2 - masz to.
Intonacja opadająca pełna.
Znacznik na skali czasu rozmieszczono co 30 ms.

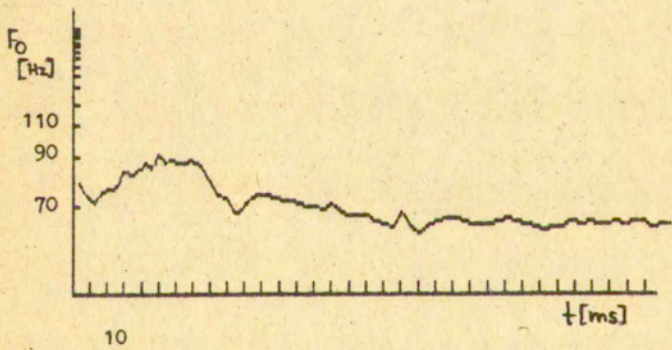
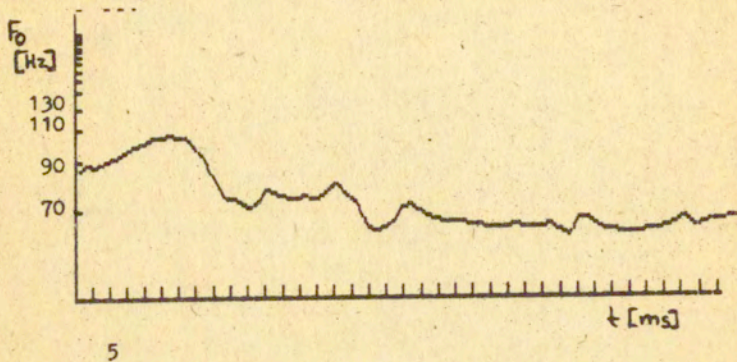


3

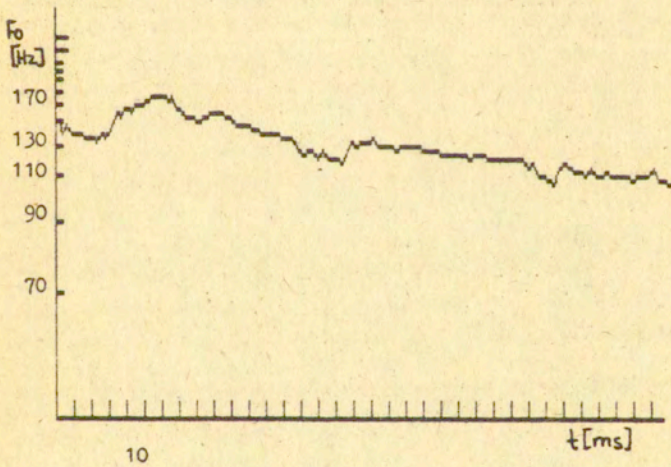
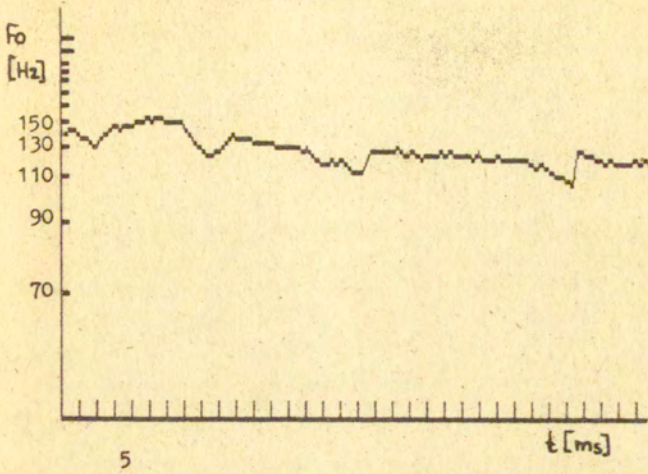


4

Rys. 13. Wypowiedzi wzorcowe: 3 - masz to tu, 4 - masz to tutaj. Intonacja pełna opadająca. Znacznik na skali czasu rozmieszczono co 30 ms.



Rys. 14. Przebiegi parametru F_0 w wypowiedziach "dawałby mi go"
Intonacja niska opadająca i wysoka opadająca. Głos WI.
Znacznik na skali czasu rozmieszczono co 30 ms.



Rys. 15. Przebiegi parametru F_0 w wypowiedziach "dawałby mi go"
Intonacja niska opadająca i wysoka opadająca. Głos HK.
Znacznik na skali czasu rozmieszczono co 30 ms.

BIBLIOGRAFIA

- ARMSTRONG, L. E. & JONES, D. (1951): *The Phonetics of French*. G. Bell & Sons Ltd., London.
- & WARD, I. C. (1944): *A Handbook of English Intonation*, 2nd ed., Heffner & Sons, Cambridge.
- BOLINGER, D. (1965): *Intonation and its Parts*, E. Arnold, Victoria, Australia.
- BOYANUS, S. C. (1944): *A Manual of Russian Pronunciation*, 2nd ed. Sidgwick & Jackson Ltd, London.
- CHAPALLAZ, M. (1964): Intonation of questions in Italian, [w:] *In Honor of Daniel Jones* (D. Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott, J. L. M. Trim, eds.), 374-383, Longman, London.
- COUSTENOBLE, H. & ARMSTRONG, L. (1934): *Studies in French Intonation*, Heffer, Cambridge.
- COWAN, J. M. (1962): Graphical representation of perceived pitch in speech, *Proceedings of the 4th Intern. Congress of Phonetic Sciences*, Mouton, The Hague, 567-570.
- CRYSTAL D. & DAVY D. (1969) *Investigating English Style*, Longman, London.
- CRUTTENDEN, A. (1986): *Intonation*, Cambridge Univ. Press, Cambridge.
- CRYSTAL, D. (1969): *Prosodic Systems and Intonation in English*, Cambridge University Press, Cambridge.
- von ESSEN, W. (1966): *Allgemeine und angewandte Phonetik*, Akademie-Verlag, Berlin.
- FAURE, G. (1962): *Recherches sur les caractères et le rôle des éléments musicaux dans la prononciation anglaise*, Didier, Paris.
- FÓNAGY, J. & MAGDICS, K. (1963): Emotional patterns in intonation and music, *Zeitschrift f. Phonetik, Sprachwissenschaft u. Kommunikationsforschung*, 16, 293-313.
- HALLIDAY, M. A. K. (1970): *A course in Spoken English: Intonation*, Oxford University Press, London.
- HESS, W. J. (1982): Algorithms and devices for pitch determination of speech signals, *Phonetica*, 39, 219-240.
- HESS, W. J. (1983): *Pitch Determination of Speech Signals*, Springer Verlag, Berlin.
- HILL, L. A. (1965): *Stress and Intonation Step by Step. Work-*

book, Oxford Univ. Press, London.

- JASSEM, W. (1952): *Intonation of Conversational English*, Prace Wrocławskiego Tow. Nauk. Ser. A, Nr 45, Wrocław.
- (1962): *Akcent języka polskiego*, Zakład Nar. im. Ossolińskich, Wyd. PAN, Prace Językoznawcze nr 31, Wrocław.
- (1987a): *The Phonology of Modern English*, PWN, Warszawa.
- 1987b): *Off-line recognition of basic F₀ curves*, *European Conference on Speech Technology*, Edinburgh, 219-222.
- , Demenko, G. & Krzysko, M.: *Classification of basic fundamental frequency patterns using discriminant functions*, *Phonetica* 44, 246-257.
- JONES, D. (1909): *Intonation Curves*, Teubner, Leipzig.
- , D. (1940): *Phonetic Readings in English*, 26th ed. Winters Universitätsbuchhandlung, Heidelberg.
- (1956a) *Phonetic Readings in English*, re-set, Winter. Universitätsbuchhandlung, Heidelberg.
- (1956b): *An Outline of English Phonetics*, 8th ed., Heffer & Sons Ltd., Cambridge.
- & WARD, D. (1969): *The Phonetics of Russian*, Cambridge Univ. Press, Cambridge.
- KENNING, M.-M. (1983): *The tones of English and French*, *Journ. of the Intern. Phon. Association*, 13, 32-48.
- KINGDON, R. (1958): *The Groundwork of English Intonation*, Longmans, Green & Co., London.
- KLINGHARDT, H. & OLBRICH, P. (1925) *Französische Intonationsübungen*. Verl. Quelle & Meyer, Leipzig (1. wyd. 1911).
- LEWIS, J.W. ((1977): *People Speaking*, Oxford Univ. Press & Cornelsen, Berlin.
- LIKEN, K. (1969): *Phonetik und Phonologie des heutigen Italienisch*, Akademie Verlag, Berlin.
- O'CONNOR, J.D. & ARNOLD, G.F. (1973): *Intonation of Colloquial English*, 2nd ed., Longman, London.
- PALMER, H.E. (1922): *English Intonation with Systematic Exercises*, Heffer & Sons Ltd. (5. wyd. 1935).
- PIKE, K.L. (1945): *The Intonation of American English*, 8th ed., Univ. Of Michigan Press, Ann Arbor.
- (1972) *General characteristics of Intonation [w:] Intonation* (D. Bolinger, ed.) Penguin Books Ltd,

Universität zu Köln, Bericht Nr. 5, Köln.

- STIEFFEN-BATOG, M. (1986): Versuch einer strukturellen Analyse der polnischen Aussagemelodie, *Zeitschrift f. Phonetik, Sprachwiss. u. Kommunikationsforschung*, 19, 397-440.
- SURMANOWICZ-DEMENKO, G. (1987): Mathematical aspects of the classification of basic pitch patterns, *European Conference on Speech Technology*, Edinburgh, 301-304.
- TRIM, J.L.M. (1984): Tonic stress-marks for German, [w:] *Honor of Daniel Jones* (P.B. Fry, P.A.D. MacCarthy, N.C. Scott, & J.L.M. Trim, eds.) Longmans, London, 374-383.

